MAP ESTIMATES

* What if we just want point estimates (like Maximum-likelihood) but with 'some' benefits of prior.

$$\hat{\theta}_{MAP} \leftarrow \underset{\theta}{\operatorname{argmax}} \log P(\theta|x)$$

$$\log P(\theta|x) = \log P(x|\theta) + \log P(\theta) - \log P(x)$$

$$\underset{\theta}{\operatorname{argmax}} \log P(\theta|x) = \underset{\theta}{\operatorname{argmax}} \underbrace{\left[\log P(x|\theta) + \log P(\theta)\right]}$$

* for conjugate priors, identical to ML estimate with pseudo-observations
* pseudo-observations serve as regularization parameters.

for example :- for multinomial likelihood & dirichlet prior, (estimation problem in Section 2-a)

$$\log P(\theta \mid x) = \sum_{k=1}^{K} \left(\sum_{n=1}^{N} x_{n,k} + \alpha_k - 1 \right) \log \theta_k$$

maximization w.r.t θ_k

$$\Longrightarrow \theta_k \propto \sum_{n=1}^{N} x_{n,k} + \underbrace{\alpha_k - 1}_{\text{pseudo-observations}}$$

* Extension to doing ML by Expectation-Maximization is similar.

* We'll see more examples in our generalization to exponential family.